数理解析研究所講究録
第 1752 巻 2011 年 113-131

Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test

九州大学 マス・フォア・インダストリ研究所 増田 弘毅 (Hiroki Masuda)
Institute of Mathematics for Industry, Kyushu University

Abstract

We consider a family of ergodic Lévy-driven stochastic differential equations observed at high-frequency discrete sampling points, where we do not suppose a specific form of the driving Lévy measure, while the coefficients are known except for finite-dimensional parameters. Our aim is twofold: first, we derive the first-order asymptotic behavior of an M-estimator based on the approximate quadratic martingale estimating function; second, as an application of the estimator obtained, we derive consistent and asymptotically distribution-free test statistics for the normality of the driving Lévy process, based on the self-normalized partial sums of the Euler-type residuals. This paper is a slightly revised version of author's preprint [14].

1 Introduction

In this paper we are concerned with the univariate Stochastic Differential Equation (SDE)

$$\begin{cases} dX_t = a(X_t,\alpha)dt + b(X_{t-},\beta)dZ_t, \\ X_0 = x_0 \end{cases} \tag{1}$$

where a and b are $\mathbb{R}$-valued functions, which are known except for the finite-dimensional parameters α and β, respectively, and Z is a univariate Lévy process. The Lévy driven SDE (1) can be regarded as an extended-noise version of the diffusion process, for which Z is a standard Wiener process. Suppose that we have a discrete-time sample

$$(X_{t_0}, X_{t_1}, \dots, X_{t_n})$$

with regular sampling points $t_i = t_i^n = ih_n$, where $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. Our main objective is twofold.

1. First, based on the available data we want to estimate the unknown parameter $\theta := (\alpha, \beta)$ under the ergodicity. It is common knowledge that the exact likelihood estimation is infeasible in general, since most often we cannot write down the transition density in an explicit form while its existence is easy to verify. Nevertheless, for the diffusion case the issue has been solved under some appropriate conditions, and there exist several explicit contrast functions. See Kessler [5] and Yoshida [25, 26] as well as the references therein: in order to derive an explicit contrast function, they employed a discretized version of the continuous-observation likelihood process, or a local-Gauss approximation of the transition density. In this case, it is well known in the literature that the quasi-likelihood type contrast function, for which only conditional mean and variances of data increments do matter, lead to an asymptotically efficient estimator.

   On the other hand, the issue has not been addressed enough in the presence of (possibly infinite-variation) jumps. Among many possibilities, in this paper we focus on the (Gaussian) quasi-likelihood type contrast function. Different from the diffusion case, it does not produce asymptotically efficient estimator in the presence of jumps. Nevertheless, it has at least two important advantages: the contrast function to be optimized is explicit; and estimation procedure is robust to modelling Lévy measure, which we actually do not need to specify.

2. Second, as an application of the estimator thus obtained, we consider testing whether or not Z is Gaussian. Under the nontriviality of Z, this can be stated as follows:

$$\mathcal{H}_0 : Z \text{ is a standard Wiener process.}$$
$$\mathcal{H}_1 : \text{Not } \mathcal{H}_0.$$

The alternative $\mathcal{H}_1$ just means that $\nu(\mathbb{R}\backslash\{0\}) \in (0, \infty]$; the Gaussian part of Z may or may not be null under $\mathcal{H}_1$. We are concerned here with formulation of a Jarque-Bera-type test for a discretely observed Markov process defined as a solution process to a stochastic differential equation. Our focus is put rather on seemingly diffusion-like process, than on diffusion plus rare and large jumps. Pure-jump Lévy process with many small jumps may be approximated in distribution by a Wiener process. Nevertheless, our test statistics can simply and successfully detect the non-Gaussianity under high frequency sampling.

In the light of prior literatures, when we apply a parametric diffusion model to high-frequency data, we blindly utilize the optimal rate $\sqrt{n}$ of the diffusion parameter, faster than the optimal rate $\sqrt{nh_n}$ of the drift parameter. It is well known that the (approximate) quadratic estimating function leads to the optimal rates in case of diffusions. However, this is *not* the case if Z admits any nontrivial jump part, even if Z is "distributionally" close to a standard Wiener process, such as the normal inverse Gaussian Lévy process such that $\mathcal{L}(Z_1) = NIG(\alpha, 0, \delta, 0)$ with large α and δ. One should pay attention to this hazardous nature in practice. This point justifies performing a normality test for Z.

This paper is organized as follows. Section 2 describe our target model and assumptions. In Section 3, we provide the asymptotic behavior of the Gaussian-reference quasi-likelihood estimator, which can be seen as the estimator based on an approximate quadratic martingale estimating function. As an application of the result of Section 3, in Section 4 we presents our noise normality test: an earlier attempts in this direction has been made by Lee and Masuda [10] (see also Kulperger and Yu [8] and Yu [28] in the context of time series analysis).

2 Lévy-driven SDE with ergodic property

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, P)$ be a complete filtered probability space endowed with a nontrivial centered Lévy process $Z = (Z_t)_{t\in\mathbb{R}_+}$ admitting an Lévy-Itô decomposition

$$Z_t = \sigma w_t + \int_0^t \int z\tilde{\mu}(ds, dz), \tag{2}$$

where $\sigma \geq 0$, w is a univariate standard Wiener process, and $\tilde{\mu}(ds, dz) = \mu(ds, dz) - ds\nu(dz)$ with a Poisson random measure on $\mathbb{R}\backslash\{0\}$ having Lévy measure ν (we refer to the monograph Sato [17] for a detailed description of Lévy processes). On this space, we consider a solution $X = (X_t)_{t\in\mathbb{R}_+}$ to the SDE (1), where $\alpha \in \Theta_\alpha \subset \mathbb{R}^{p_\alpha}$ and $\beta \in \Theta_\beta \subset \mathbb{R}^{p_\beta}$, with Θ_α and Θ_β being convex domains. We write $\Theta = \Theta_\alpha \times \Theta_\beta$ for the parameter space of $\theta = (\alpha, \beta)$, and denote by $p := p_\alpha + p_\beta$ the dimension of the all unknown parameters involved. The initial variable X_0 is $\mathcal{F}_0$-measurable and independent of Z, and the coefficients $a : \mathbb{R} \times \Theta_\alpha \to \mathbb{R}$ and $b : \mathbb{R} \times \Theta_\beta \to \mathbb{R}$ are known measurable functions. We denote by $\theta_0 = (\alpha_0, \beta_0) \in \Theta$ the true value of θ, supposed to exist. We observe X at time points $t_i^n = ih_n$, so that available data is $X_0, X_{h_n}, \ldots, X_{nh_n}$. Throughout this paper we work under the "rapidly increasing experimental design":

$$h_n \to 0, \quad T_n := nh_n \to \infty, \quad nh_n^2 \to 0. \tag{3}$$

Here and in what follows asymptotics are taken for $n \to \infty$ unless otherwise mentioned.

Our underlying statistical model is the parametric family $(P_\theta)_{\theta\in\Theta}$, where P_θ stands for the image measure of X associated with $\theta \in \Theta$. For simplicity, we abbreviate P_{θ_0} as P_0.

We use the following notation.

- C denotes a positive generic constant, which may vary from line to line.
- $\Delta_i X := X_{t_i^n} - X_{t_{i-1}^n}$.
- $f_{i-1}(v) := f(X_{t_{i-1}}, v)$ for a variable $v \in V$ and a measurable function f on $\mathbb{R} \times V$.

- $\to^p$ and $\to^d$ denotes the convergence in P_0 probability and the convergence in law under P_0, respectively.

We impose the following regularity conditions on the coefficients.

Assumption 2.1. *Either one of the following two holds true.*

1. *(Bounded smooth coefficients plus uniformly nondegenerate $b(x,\beta)$)*
   - $a \in C^{\infty,2}(\mathbb{R}\times\Theta_\alpha)$ *and* $b \in C^{\infty,2}(\mathbb{R}\times\Theta_\beta)$.
   - $\sup_{(x,\theta)\in\mathbb{R}\times\Theta}\{|\partial_x^j\partial_\alpha^k a(x,\alpha)| \vee |\partial_x^j\partial_\beta^k b(x,\beta)|\} < \infty$ *for each* $j \in \mathbb{Z}_+$ *and* $k \in \{0,1,2\}$.
   - $\inf_{(x,\beta)\in\mathbb{R}\times\Theta_\beta}|b(x,\beta)| > 0$.
2. *(Globally Lipschitz smooth coefficients plus nondegenerate $b(x,\beta)$)*
   - $a \in C^{\infty,2}(\mathbb{R}\times\Theta_\alpha)$ *and* $b \in C^{\infty,2}(\mathbb{R}\times\Theta_\beta)$.
   - $\sup_{(x,\theta)\in\mathbb{R}\times\Theta}\{|\partial_x a(x,\alpha)| \vee |\partial_x b(x,\beta)|\} < \infty$.
   - $\sup_{(x,\theta)\in\mathbb{R}\times\Theta}\{|\partial_x^j\partial_\alpha^k a(x,\alpha)| \vee |\partial_x^j\partial_\beta^k b(x,\beta)|\} \le C(1+|x|)^C$ *for each* $j \ge 2$ *and* $k \in \{0,1,2\}$.
   - $\sup_{(x,\beta)\in\mathbb{R}\times\Theta_\beta}|b(x,\beta)|^{-1} \le C(1+|x|)^C$.

Under Assumption 2.1, the SDE (1) admits a unique strong solution without reference to the concrete structure of Z; see, e.g., Protter [15] for details. The reason why we separate the cases 1 and 2 in Assumption 2.1 is related to validity of the (conditional-) moment estimates we repeatedly need in the proofs. In prior literatures concerning estimation of discretely observed SDEs, it is most often assumed that $\sup_t E[|X_t|^q] < \infty$ for every $q > 0$ (see, among others, Kessler [5], Masuda [11], Shimizu and Yoshida [18], Sørensen [19]). Specifically, this is required since it is also supposed that the coefficients together with their partial derivatives are of (at most) polynomial growth uniformly in both of state variable and parameters, while in most cases the positive-order derivatives are uniformly bounded. The L^p-boundedness of X often rules out some important classes of X, such as Langevin diffusions with heavy-tailed invariant density g, described by

$$dX_t = \left(\frac{1}{2}b^2(X_t)\partial\log g(X_t) + b(X_t)\partial b(X_t)\right)dt + b(X_t)dw_t, \tag{4}$$

with b being bounded; in this case X may exhibits a kind of long-range dependence in the sense that the mixing (absolutely-regular) rate is at most subexponential or polynomial (see, e.g., Roberts and Tweedie [16, Theorem 2.4], Stramer and Tweedie [20, Theorem 3.1(i)], and Veretennikov [22, 23]). A little bit surprisingly, this point does not seem to have been specified in the literature. Therefore, we incorporate it as the first one of Assumption 2.1; in case of the second one of Assumption 2.1, we will additionally assume $\sup_t E[|X_t|^q] < \infty$ for every $q > 0$. Clearly, minimal q depends on the growth orders of a and b as well as of their partial derivatives, and on the tail of ν. Here we do not specify this intermediate case for the sake of simplicity.

The next condition is the ergodicity of X.

Assumption 2.2. *There exists a unique invariant distribution π_0 (depending on θ_0) such that we have the law of large numbers:*

$$\frac{1}{T_n}\int_0^{T_n} f(X_t)dt \to^p \int f(x)\pi_0(dx) \tag{5}$$

as soon as $f \in L^1(\pi_0)$.

Several sets of sufficient conditions to verify Assumption 2.2 are available: see, e.g., the references of Masuda [12] and Kulik [6, Appendix A.1]. Especially for general diffusions with compound-Poisson jumps (i.e., $\nu(\mathbb{R}) < \infty$), quite simple conditions are given by, e.g., Masuda [13].

Finally, we impose moment conditions on the driving Lévy process Z of the form (2).

Assumption 2.3. *There exists an integer $q > (p \vee 8)$ such that $E[|Z_t|^q] < \infty$, and $E[Z_t^2] = t$ for each $t \in \mathbb{R}_+$.*

We have imposed the condition "$q > (p \vee 8)$" just for providing a consistent estimator of the asymptotic covariance matrix in Theorem 3.4 in a direct manner. This condition might be relaxed by taking a closer look at the series of estimates in the proofs.

3 Approximate quadratic martingale estimating function

Here we derive an asymptotic normality result concerning the M-estimator of θ based on an approximate quadratic martingale function, which essentially amounts to the local-Gauss transition fitting. How to choose an estimating function could be a lot of things, possibly depending on specific structure of the jump part of the driving Lévy process. Nevertheless, as seen in the quasi-likelihood analysis in the time-series literature, the quadratic type estimating function is expected to possess high versatility, thereby being of "practical" use.

Let $A : \mathbb{R} \times \Theta_\alpha \to \mathbb{R}^{p_\alpha}$ and $B : \mathbb{R} \times \Theta_\beta \to \mathbb{R}^{p_\beta}$ be measurable functions. We define a class of random functions $Q_n : \Theta \to \mathbb{R}^p$ by

$$Q_n(\theta) = \begin{pmatrix} Q_n^1(\theta) \\ Q_n^2(\theta) \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} A_{i-1}(\theta)(\Delta_i X - h_n a_{i-1}(\alpha)) \\ B_{i-1}(\theta)\{(\Delta_i X - h_n a_{i-1}(\alpha))^2 - h_n b_{i-1}(\beta)^2\} \end{pmatrix}. \tag{6}$$

We target at estimators $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ of θ such that $Q_n(\hat{\theta}_n) = 0$ (for n large enough). Note that we are in a semiparametric framework in the sense that we only impose moment structures about the driving process Z. Needless to say, the estimator treated here cannot be asymptotically efficient in the presence of jumps. Nevertheless, it has an obvious advantage, robustness to the specification of Z's Lévy measure.

Remark 3.1. *The estimating function $Q_n(\theta)$ stems from the leading-term approximation of the genuine quadratic estimating function*

$$Q_n^*(\theta) = \sum_{i=1}^n \begin{pmatrix} A_{i-1}(\theta)(X_{t_i} - E_\theta^{i-1}[X_{t_i}]) \\ B_{i-1}(\theta)\{(X_{t_i} - E_\theta^{i-1}[X_{t_i}])^2 - E_\theta^{i-1}[(X_{t_i} - E_\theta^{i-1}[X_{t_i}])^2]\} \end{pmatrix},$$

where $E_\theta^{i-1}[\cdot] := E_\theta[\cdot|\mathcal{F}_{t_{i-1}}]$. This can be explicit if, for example, $a(x,\alpha)$ is linear in x. We here do not pay special attention to this case, however, it is obvious from the proof of Theorem 3.4 below that asymptotic behavior of the M-estimator associated with $Q_n^*(\theta)$ can be obtained under appropriate conditions in a manner similar to the case of $Q_n(\theta)$. Sørensen [19] studied general approximate martingale function for diffusion processes, including the efficiency result in Kessler [5].

We introduce some additional assumptions.

Assumption 3.2. *Either one of the following holds true.*

1. *We have Assumptions 2.1.1, 2.2, and 2.3. Additionally,*
   - $A(x,\cdot) \in C^2(\Theta)$ *and* $B(x,\cdot) \in C^2(\Theta)$ *for every* $x \in \mathbb{R}$,
   - $\sup_{(x,\theta)\in\mathbb{R}\times\Theta}\{|\partial_\theta^k A(x,\theta)| \vee |\partial_\theta^k B(x,\theta)|\} < \infty$ *for* $k \in \{0,1,2\}$.
2. *We have Assumptions 2.1.2, 2.2, and 2.3. Additionally,*
   - $A(x,\cdot) \in C^2(\Theta)$ *and* $B(x,\cdot) \in C^2(\Theta)$ *for every* $x \in \mathbb{R}$,
   - $\sup_{\theta\in\Theta}\{|\partial_\theta^k A(x,\theta)| \vee |\partial_\theta^k B(x,\theta)|\} \le C(1+|x|)^C$ *for each* $k \in \{0,1,2\}$,
   - $\sup_{t\in\mathbb{R}_+} E_0[|X_t|^q] < \infty$ *for every* $q > 0$.

Typically, the last condition in Assumption 3.2.2 follows from a kind of Foster-Lyapunov criteria, which is usually easy to verify: see Masuda [12, 13] for details.

We write $\pi_0(f(\cdot,\theta)) = \int f(x,\theta)\pi_0(dx)$ for a function on $\mathbb{R}\times\Theta$, and often abbreviate $\pi_0(f(\cdot,\theta_0))$ to $\pi_0(f)$. We now impose the identifiability condition:

Assumption 3.3. *The matrices $\pi_0(A\partial_\alpha^\top a)$ and $\pi_0(B\partial_\beta^\top b^2)$ are nonsingular, and the identity*

$$|\pi_0(A(\cdot,\theta)\{a(\cdot,\alpha_0) - a(\cdot,\alpha)\})| + |\pi_0(B(\cdot,\theta)\{b^2(\cdot,\beta_0) - b^2(\cdot,\beta)\})| = 0$$

holds true iff $\theta = \theta_0$.

From now on, we write

$$\nu_k = \int z^k \nu(dz)$$

for $k \geq 3$, $k \in \mathbb{N}$; of course, ν_k is nothing but the kth cumulant of $\mathcal{L}(Z_1)$, hence we have $\nu_3 = E[Z_1^3]$ and $\nu_4 = E[Z_1^4] - 3$ under Assumption 2.3. Under the aforementioned assumptions, the following $p \times p$ symmetric matrices are well-defined:

$$V_0' = \mathrm{diag}(V_{11}, 2V_{22}), \quad V_0'' = \begin{pmatrix} V_{11} & \nu_3 V_{12} \\ \text{Sym.} & \nu_4 V_{22} \end{pmatrix},$$

where

$$\begin{aligned} V_{11} &= \{\pi_0(A\partial_\alpha^\top a)^\top\}^{-1}\pi_0(A^{\otimes 2}b^2)\pi_0(A\partial_\alpha^\top a)^{-1}, \\ V_{12} &= \{\pi_0(A\partial_\alpha^\top a)^\top\}^{-1}\pi_0(AB^\top b^3)\pi_0(B\partial_\beta^\top b^2)^{-1}, \\ V_{22} &= \{\pi_0(B\partial_\beta^\top b^2)^\top\}^{-1}\pi_0(B^{\otimes 2}b^4)\pi_0(B\partial_\beta^\top b^2)^{-1}. \end{aligned}$$

For a function $f(x, \theta)$ on $\mathbb{R} \times \Theta$, we write

$$\hat{S}_n(f) = \frac{1}{n}\sum_{i=1}^n f(X_{t_{i-1}}, \hat{\theta}_n).$$

Here is the main claim of this section.

Theorem 3.4. *Suppose Assumptions 2.3, 3.2, and 3.3. Then, there exists a $\sigma(X_{t_i^n} : i \leq n)$-measurable $\mathbb{R}^p$-valued sequence $\hat{\theta}_n$ such that $P_0[Q_n(\hat{\theta}_n) = 0] \to 1$, and any such sequence fulfils $\hat{\theta}_n \to^p \theta_0$. Moreover:*

$$\text{if } \nu(\mathbb{R}) = 0, \text{ then } \left(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0)\right) \to^d \mathcal{N}_p(0, V_0'); \tag{7}$$

$$\text{if } \nu(\mathbb{R}) > 0, \text{ then } \sqrt{T_n}(\hat{\theta}_n - \theta_0) \to^d \mathcal{N}_p(0, V_0''). \tag{8}$$

Consistent estimators of the asymptotic variances can be given through through the following: in case of (7),

$$\begin{aligned} \hat{V}_{11,n} &= \{\hat{S}_n(A\partial_\alpha^\top a)^\top\}^{-1}\hat{S}_n(A^{\otimes 2}b^2)\hat{S}_n(A\partial_\alpha^\top a)^{-1} \\ \hat{V}_{22,n} &= \{\hat{S}_n(B\partial_\beta^\top b^2)^\top\}^{-1}\hat{S}_n(B^{\otimes 2}b^4)\hat{S}_n(A\partial_\alpha^\top a)^{-1}; \end{aligned}$$

and in case of (8), $\hat{V}_{11,n}$ is the same as above, while

$$\widehat{\nu_3 V_{12,n}} = \{\hat{S}_n(A\partial_\alpha^\top a)^\top\}^{-1}\left(\frac{1}{T_n}\sum_{i=1}^n AB^\top(X_{t_{i-1}}, \hat{\theta}_n)\{\Delta_i X - h_n a_{i-1}(\hat{\alpha}_n)\}^3\right)\hat{S}_n(B\partial_\beta^\top b^2)^{-1},$$

$$\widehat{\nu_4 V_{22,n}} = \{\hat{S}_n(B\partial_\beta^\top b^2)^\top\}^{-1}\left(\frac{1}{T_n}\sum_{i=1}^n B^{\otimes 2}(X_{t_{i-1}}, \hat{\theta}_n)\{\Delta_i X - h_n a_{i-1}(\hat{\alpha}_n)\}^4\right)\hat{S}_n(B\partial_\beta^\top b^2)^{-1}.$$

The diffusion case (7) is well known and formally not new, and we know the optimal choices of A and B leading to the asymptotic efficiency of the corresponding $\hat{\theta}_n$. Nevertheless, the author could not find any literature that specify the claim under possibly heavy-tailed X, which may occur under Assumption 2.1.1; indeed, this point does seem to have been ignored so far.

The non-null μ case (8) is new, but the asymptotic efficiency of $\hat{\theta}_n$ is no longer valid. The efficiency issue is left as an important (and quite intricate) open problem.

Remark 3.5. *We can provide, at least formally, a similar result also in case where $nh^r \to 0$, $r \geq 2$, $r \in \mathbb{N}$, by using higher-order Itô-Taylor expansions for the conditional mean and variance. However, we then inevitably need to know moment structure of order $k \geq 3$ about the Lévy measure ν. This implies that our estimating procedure loses the merit of the robustness to modelling the essentially "infinite-dimensional" nuisance parameter ν. Furthermore, even if we could know the moment structure, the resulting estimating function then looks much more complicated and its optimization becomes harder, thereby diminishing its usefulness in practice. See also Remark 5.4.*

Remark 3.6. *Let $\nu_3 = 0$ and $p_\alpha = p_\beta = 1$. Then, in view of Schwarz's inequality, it is easy to see that*

$$A(x,\theta) = \frac{\partial_\alpha a(x,\alpha)}{b(x,\beta)^2} \quad \text{and} \quad B(x,\theta) = \frac{\partial_\beta [b(x,\alpha)^2]}{b(x,\beta)^4}$$

is an asymptotically optimal choice: interestingly, this optimal choice is the same as in the diffusion case.

Remark 3.7. *We could also deduce a "mighty convergence" result for multivariate X, meaning that in addition to the weak convergence of the normalized $\hat{\theta}_n$ we could also have their $L^q(P_0)$-boundedness. This will be reported elsewhere.*

4 Normality test for the driving noise

In the previous section, we have seen that the approximate quadratic estimating function (6) can be used regardless of the presence of jumps. However, within our underlying model (1) with Z unknown, we beforehand do not know which of a diffusion or a process with jumps is more appropriate to given data. This underlines the importance of testing normality of Z.

In this section, we are concerned with testing the normality of Z against presence of any nontrivial jump component:

$$\mathcal{H}_0 : Z \text{ is a standard Wiener process.}$$
$$\mathcal{H}_1 : \text{Not } \mathcal{H}_0.$$

Since we presuppose that Z is nontrivial, the alternative hypothesis specifically means that $\nu(\mathbb{R}) \in (0,\infty]$. The Gaussian component of Z may or may not degenerate under $\mathcal{H}_1$.

For convenience, we remark the prototype of the *Jarque-Bera test for normality*. Let $\xi_1, \xi_2, \ldots$ be a sequence of i.i.d. random variables, and for $k \geq$ let

$$\hat{z}_n^{(k)} := \frac{n^{-1}\sum_{i=1}^n (\xi_i - \bar{\xi}_n)^k}{\{n^{-1}\sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2\}^{k/2}},$$

where $\bar{\xi}_n := n^{-1}\sum_{i=1}^n \xi_i$. Then the Jarque-Bera statistics

$$\mathcal{J}_n := n\left[\frac{(\sqrt{n}\hat{z}_n^{(3)})^2}{6} + \frac{\{\sqrt{n}(\hat{z}_n^{(4)} - 3)\}^2}{24}\right]$$

weakly tends to $\chi^2(2)$ as soon as the law of ξ_1 is normal, providing a simple procedure of testing normality. This test is based on the fact: if $\sqrt{n}(V_n - V) \to^d \mathcal{N}_p(0,\Sigma)$ for some constants $V \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^p \otimes \mathbb{R}^p$, then it follows from the continuous mapping theorem that $\Sigma_n^{-1}[\{\sqrt{n}(V_n - V)\}^{\otimes 2}] \to^d \chi^2(p)$, where $\Sigma_n \to^p \Sigma$.

We consistently use the notation introduced in Section 3. Let

$$\epsilon_{ni}(\theta) := \frac{\Delta_i X - h_n a_{i-1}(\alpha)}{b_{i-1}(\beta)\sqrt{h_n}} \tag{9}$$

for $i \leq n$ and $\theta \in \Theta$, and $\hat{\epsilon}_{ni} := \epsilon_{ni}(\hat{\theta}_n)$, where $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ is the estimator introduced in Section 3. Write $\bar{\hat{\epsilon}}_n = n^{-1}\sum_{i=1}^n \hat{\epsilon}_{ni}$, $\hat{\Psi}_n^{(k)} = n^{-1}\sum_{i=1}^n (\hat{\epsilon}_{ni} - \bar{\hat{\epsilon}}_n)^k$, and $\Phi_n^{(k)} = \hat{\Psi}_n^{(k)}(\hat{\Psi}_n^{(2)})^{-k/2}$. Our test statistics is then defined to be

$$\mathcal{T}_n = \frac{n}{6}\left\{\hat{\Phi}_n^{(3)} - \frac{3\sqrt{h_n}}{n}\sum_{i=1}^n \partial_x b(X_{t_{i-1}}, \hat{\beta}_n)\right\}^2 + \frac{n}{24}(\hat{\Phi}_n^{(4)} - 3)^2. \tag{10}$$

Given a value of $\hat{\theta}_n$, it is straightforward to evaluate $\mathcal{T}_n$. This simple test statistics turns out to be asymptotically distribution-free and consistent.

Theorem 4.1. *Suppose that the conditions of Theorem 3.4 are in force. Then, we have the following:*

- *$\mathcal{T}_n \to^d \chi^2(2)$ under $\mathcal{H}_0$;*
- *$P_0[\mathcal{T}_n > K] \to 1$ for every $K > 0$ under $\mathcal{H}_1$.*

Here are some remarks on Theorem 4.1.

Remark 4.2. *Theorem 4.1 extends Lee and Masuda [10], where we targeted at constant $b(x,\beta)$. Although the construction of our test statistics (the Jarque-Bera methodology) is somewhat analogous, we need to be careful in handling the dispersion term $b(x,\beta)$. The test statistics $\mathcal{T}_n$ are different from what is proposed in Lee and Masuda [10]: in the present setting, we have a bias-correction term in the sample-skewness part. We note that $\mathcal{T}_n$ reduces to that of Lee and Masuda [10] in cases where $b(x,\beta)$ is constant. It is possible to construct $\mathcal{T}_n$ based on higher-order self-normalized partial sums of residuals, however, performance of $\mathcal{T}_n$ under $\mathcal{H}_0$ may then deteriorate since higher-order sample moments appear in the statistics, while power can be gained due to resulting bigger variance of $\mathcal{T}_n$.*

Remark 4.3. *In the proof of Theorem 4.1, we have to specify the behaviors of $\hat{\theta}_n$ to some extent under both $\mathcal{H}_0$ and $\mathcal{H}_1$. In our case, it is crucial that $\hat{\theta}_n$ is rate-optimal under $\mathcal{H}_0$ and at the same time $\sqrt{T_n}$-consistent under $\mathcal{H}_1$. If instead of $\hat{\theta}_n$ we adopt some estimator $\hat{\theta}_n$, which is only $\sqrt{T_n}$-consistent even under $\mathcal{H}_0$, then, from the proof of Theorem 4.1, $\mathcal{T}_n$ would not work: more precisely, the required rate of $\hat{\theta}_n$ mentioned above are enough to conclude that the effect of plugging-in $\hat{\theta}_n$ vanishes in the leading term of $\mathcal{T}_n$ (see the proof of Theorem 4.1 for details). In this sense, the approximate quadratic estimating function is quite natural to adopt in our framework, for it produce possibly the simplest rate-optimal estimator under $\mathcal{H}_0$, and at the same time, $\sqrt{T_n}$-consistency under $\mathcal{H}_1$.*

Remark 4.4. *From the proof of the consistency of the test (Section 5.3.2), it is clear that we may adopt*

$$\mathcal{T}_n' := \frac{n}{6}\left\{\hat{\Phi}_n^{(3)} - \frac{3\sqrt{h_n}}{n}\sum_{i=1}^{n}\partial_x b(X_{t_{i-1}},\hat{\beta}_n)\right\}^2$$

instead of $\mathcal{T}_n$: then, $\mathcal{T}_n' \to^d \chi^2(1)$ under $\mathcal{H}_0$ and the consistency remains valid. That is to say, in order to derive the consistency against the "full" alternative $\mathcal{H}_1$ (i.e., presence of "any" nontrivial jump component), it is enough to look at the sample skewness solely. This is a sharp contrast to the classical i.i.d.-sample case; one can consult Henze [2] for an extensive review of the consistency issue in testing (possibly multivariate) normality.

Remark 4.5. *For computing $\hat{\Phi}_n^{(k)}$, the factor $\sqrt{h_n}$ in the denominator of (9) is clearly redundant. However, we keep it for convenience of references in the proof.*

5 Proofs

Throughout this section we use the following notation.

- $R_1(x,\theta)$ (resp. $R_2(x,\theta)$) denotes generic functions on $\mathbb{R}\times\Theta$ such that $\sup_{x,\theta}|R_1(x,\theta)| < \infty$ (resp. $\sup_\theta |R_2(x,\theta)| \le C(1+|x|)^C$). Dimensions of $R_j(x,\theta)$ vary depending on the context.
- ρ_k stands for the kth moment of the standard normal distribution.
- For random sequences (x_n) and (y_n), $x_n \lesssim y_n$ means that $x_n \le C y_n$ a.s. for every n large enough.
- $\zeta_{ni}' := \Delta_i X - h_n a_{i-1}(\alpha_0)$ and $\zeta_{ni}'' := \{\Delta_i X - h_n a_{i-1}(\alpha_0)\}^2 - h_n b_{i-1}(\beta_0)^2$.

5.1 Leading terms of Itô-Taylor expansions

We begin with fundamental facts concerning conditional size of X's increments.

Lemma 5.1. *Let $g(x,\theta) := |a(x,\alpha)| \vee |b(x,\beta)|$. Fix any $q \geq 2$ such that $E[|Z_t|^q] < \infty$. Then*

$$E_\theta^{i-1}\left[\sup_{s\in[t_{i-1},t_i]} |X_s - X_{t_{i-1}}|^q\right] \lesssim \begin{cases} h_n^{q/2} g(X_{t_{i-1}},\theta)^q, & \text{if } \nu(\mathbb{R}) = 0, \\ h_n g(X_{t_{i-1}},\theta)^q, & \text{otherwise.} \end{cases} \tag{11}$$

In particular, the left-hand side of (11) are essentially bounded if so is g.

Proof. Let $\nu(\mathbb{R}) > 0$, and $\tau_{i-1,K} := \inf\{s \geq t_{i-1} : |X_s| \geq K\}$ for $K > 0$. By means of Assumption 2.1, triangular and martingale inequalities (consult, e.g., Protter [15, Section V.11] with minor modifications), we see that $\xi_{i-1,K}(s) := E_\theta^{i-1}[\sup_{u\in[t_{i-1},s\wedge\tau_{i-1,K}]} |X_u - X_{t_{i-1}}|^q]$ fulfils $\xi_{i-1,K}(t_i) \lesssim \int_{t_{i-1}}^{t_i} \xi_{i-1,K}(s)ds + h_n g(X_{t_{i-1}},\theta)^q$, the upper bound being a.s. finite according to the definition of $\tau_{i-1,K}$. Hence the claim follows on applying Gronwall's inequality and then letting $K \to \infty$. The case where $\nu(\mathbb{R}) = 0$ is similar (see also Kessler [5, Lemma 6]). □

For a smooth function $f : \mathbb{R} \to \mathbb{R}$, the generator of X under P_θ is formally given by

$$\begin{aligned} \mathcal{A}_\theta f(x) &= a(x,\alpha)\partial f(x) + \frac{1}{2} b(x,\beta)^2 \partial^2 f(x) \\ &\quad + \int \{f(x + b(x,\beta)z) - f(x) - \partial f(x) b(x,\beta) z\}\nu(dz). \end{aligned} \tag{12}$$

Given any $m \in \mathbb{N}$, successive applications of Itô's formula yield that

$$\begin{aligned} E_\theta^{i-1}[f(X_{t_i})] &= \sum_{l=0}^{m} \frac{h_n^l}{l!} \mathcal{A}_\theta^l f(X_{t_{i-1}}) \\ &\quad + \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s_1} \cdots \int_{t_{i-1}}^{s_m} E_\theta^{i-1}\left[\mathcal{A}_\theta^{m+1} f(X_{s_{m+1}})\right] ds_{m+1} ds_m \ldots ds_1 \\ &\quad + E_\theta^{i-1}\left[M(f)_{t_{i-1}}^{t_i}\right] + \int_{t_{i-1}}^{t_i} E_\theta^{i-1}\left[M(\mathcal{A}_\theta f)_{t_{i-1}}^{s_1}\right] ds_1 \\ &\quad + \sum_{l=1}^{m-1} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s_1} \cdots \int_{t_{i-1}}^{s_l} E_\theta^{i-1}\left[M(\mathcal{A}_\theta^{l+1} f)_{t_{i-1}}^{s_{l+1}}\right] ds_{l+1} ds_l \ldots ds_1, \end{aligned} \tag{13}$$

where, for $t \geq s$,

$$M(g)_s^t := \frac{1}{2} \int_s^t \partial g(X_u) b(X_u,\beta) dw_u + \int_s^t \int \{g(X_{u-} + b(X_{u-},\beta)z) - g(X_{u-})\}\tilde{\mu}(du,dz). \tag{14}$$

(13) is valid as soon as all the appearing conditional expectations a.s. exist.

For $i \leq n$ and $k \in \mathbb{N}$, we let

$$F_{k,i}(\theta) := E_\theta^{i-1}\left[\{\Delta_i X - h_n a_{i-1}(\alpha)\}^k\right].$$

We need the leading terms of the stochastic expansions of $F_{k,i}(\theta)$ in h-power series (the so-called Itô-Taylor expansions), which play a central role in deriving the asymptotic behaviors of our estimator and test statistics.

Lemma 5.2. *Suppose $\nu(\mathbb{R}) = 0$, so that X is a diffusion process. Then, for each $i \leq n$ we have, under Assumption 2.1.1,*

$$\begin{aligned} F_{1,i}(\theta) &= h_n^2 R_1(X_{t_{i-1}},\theta), \\ F_{2,i}(\theta) &= h_n b_{i-1}(\beta)^2 + h_n^2 R_1(X_{t_{i-1}},\theta), \\ F_{3,i}(\theta) &= 3h_n^2 b(X_{t_{i-1}},\beta)^3 \partial_x b(X_{t_{i-1}},\beta) + h_n^3 R_1(X_{t_{i-1}},\theta), \end{aligned}$$

also, for $k \geq 4$,

$$F_{k,i}(\theta) = \begin{cases} h_n^{(k+1)/2} R_1(X_{t_{i-1}},\theta), & \text{for } k \text{ odd}, \\ h_n^{k/2} \rho_k b(X_{t_{i-1}},\beta)^k + h_n^{(k+2)/2} R_1(X_{t_{i-1}},\theta), & \text{for } k \text{ even}. \end{cases}$$

The same expressions except that all the R_1 are replaced by R_2 remain valid under Assumption 2.1.2.

Proof. We only prove the lemma under Assumption 2.1.1; the proof under Assumption 2.1.2 is quite similar.

Letting $f_{k,x}(y) := (y-x)^k$, we look at $E_\theta^{i-1}[f_{k,X_{t_{i-1}}}(X_{t_i})]$ by making use of (13) with f replaced by $f_{k,X_{t_{i-1}}}$. It is clear that $|\mathcal{A}_\theta^{l'} f_{k,x}(y)| \leq C(1+|y-x|^{k-1})$ for each $l' \in \mathbb{N}$. Hence, in view of (14) with the purely discontinuous local martingale part dropped out, it follows from Lemma 5.1 and Burkholder-Davis-Gundy inequality that

$$E_\theta^{i-1}\left[\sup_{s\in[t_{i-1},t_i]} \left|M(\mathcal{A}_\theta^l f_{k,X_{t_{i-1}}})_{t_{i-1}}^s\right|^2\right] \leq Ch_n.$$

Therefore, the last three terms in the right-hand side of (13) for any m vanishes as a matter of fact (see Protter [15, Theorem I.51]). Moreover, for each m, the second term in the right-hand side of (13) can be bounded by Ch_n^{m+1}. In view of (12),

$$\begin{cases} \mathcal{A}_\theta^l f_{k,x}(y) \equiv 0 \text{ for each } l \leq [k/2] & \text{if } k \in \mathbb{N} \text{ is odd,} \\ \mathcal{A}_\theta^l f_{k,x}(y) \equiv 0 \text{ for each } l \leq [k/2]-1 & \text{if } k \in \mathbb{N} \text{ is even.} \end{cases}$$

Now we deduce all the claims by easy algebra. □

In the presence of any nontrivial jump component, the leading terms of $F_{k,i}(\theta)$ for $k \geq 2$ are of $O_p(h_n)$:

Lemma 5.3. *Suppose that $\nu(\mathbb{R}) \in (0,\infty]$, $\int |z|^k \nu(dz) < \infty$ for some $k \geq 2$, with $E[Z_1] = 0$ and $E[Z_1^2] = 1$. Then, for each $i \leq n$ we have under Assumption 2.1.1,*

$$F_{k,i}(\theta) = \begin{cases} h_n^2 R_1(X_{t_{i-1}},\theta), & \text{for } k=1, \\ h_n b(X_{t_{i-1}},\beta)^2 + h_n^2 R_1(X_{t_{i-1}},\theta), & \text{for } k=2, \\ h_n \nu_k b(X_{t_{i-1}},\beta)^k + h_n^2 R_1(X_{t_{i-1}},\theta), & \text{for } k \geq 3. \end{cases} \tag{15}$$

The same expressions except that all the R_1 are replaced by R_2 remain valid under Assumption 2.1.2.

Proof. Again we only prove under Assumption 2.1.1 as the proof under Assumption 2.1.2 is similar with minor modifications.

Let $f_{k,x}(y) := (y-x)^k$ as before. We here need to look at (13) and (14) for $f = f_{k,X_{t_{i-1}}}$ and $m=1$ in more detail. Under the condition $\int |z|^k \nu(dz) < \infty$, it follows from (12) that (rather roughly)

$$\begin{aligned} |\partial(\mathcal{A}_\theta f_{k,x})(y)| &\lesssim |\partial f_{k,x}(y)| + |\partial^2 f_{k,x}(y)| + |\partial^3 f_{k,x}(y)| \\ &\quad + \left|\int \left(\int_0^1 u \int_0^1 \partial^3 f_{k,x}(y+uvb(y)z)dvdu\right) b(y)^2 z^2 \nu(dz)\right| \\ &\quad + \left|\int \left(\int_0^1 \partial^2 f_{k,x}(y+ub(y)z)du\right) b(y)\partial b(y) z^2 \nu(dz)\right| \\ &\lesssim 1 + |y-x| + \cdots + |y-x|^{k-1}, \end{aligned}$$

the upper bound remaining valid for $|\partial^2(\mathcal{A}_\theta f_{k,x})(y)|$. Hence, from the expression

$$\begin{aligned} \mathcal{A}^2 f_{k,x}(y) &= a(y)\partial(\mathcal{A}_\theta f_{k,x})(y) + \frac{1}{2} b(y)^2 \partial^2(\mathcal{A}_\theta f_{k,x})(y) \\ &\quad + \int \left\{\int_0^1 u \int_0^1 \partial^2(\mathcal{A}_\theta f_{k,x})(y+uvb(y)z)dudv\right\} b(y)^2 z^2 \nu(dz), \end{aligned}$$

we also deduce the estimate $|\mathcal{A}^2 f_{k,x}(y)| \lesssim 1 + |y-x| + \cdots + |y-x|^{k-1}$. This combined with Lemma 5.1 readily implies that, in (13): $E_\theta^{i-1}[M(\mathcal{A}_\theta^l f)_{t_{i-1}}^{s_0}] = 0$, $E_\theta^{i-1}[M(f)_{t_{i-1}}^{t_i}] = 0$, and $E_\theta^{i-1}[|\mathcal{A}^2 f_{k,x}(y)|] \lesssim 1$ (Indeed, the first two can be verified just like the proof of Lemma 5.2). Thus we obtain

$$E_\theta^{i-1}\left[f_{k,X_{t_{i-1}}}(X_{t_i})\right] = h_n \mathcal{A}_\theta f_{k,X_{t_{i-1}}}(X_{t_{i-1}}) + h_n^2 R_1(X_{t_{i-1}},\theta)$$

for each $k \geq 2$: of course, the first term in the right-hand side is $h_n a_{i-1}(\alpha)$ when $k = 1$. The claim now follows on the binomial expansion

$$\begin{aligned} F_{k,i}(\theta) &= \sum_{l=0}^{k} \binom{k}{l} \{-a_{i-1}(\alpha)\}^{k-l} h_n^{k-l} E_\theta^{i-1}\left[f_{l,X_{t_{i-1}}}(X_{t_i})\right], \\ &= h_n \mathcal{A}_\theta f_{k,X_{t_{i-1}}}(X_{t_{i-1}}) + h_n^2 R_1(X_{t_{i-1}}, \theta) \end{aligned}$$

combined with the expressions $\mathcal{A}_\theta f_{1,x}(x) = a(x,\alpha)$, $\mathcal{A}_\theta f_{2,x}(x) = b(x,\beta)^2$ (because of $E[Z_1^2] = 1$), and $\mathcal{A}_\theta f_{k,x}(x) = \nu_k b(x,\beta)^k$ for $k \geq 3$. □

Remark 5.4. *The higher-order expansions can be derived through straightforward but messy and lengthy computations; they might be practically intractable in optimization of the corresponding estimating functions. Needless to say, the principle can apply to a more general multivariate diffusions with jumps, where diffusion and jump-part coefficient functions may differ. Although we do not need them in this paper, it may be used to construct estimating functions for diffusions with jumps.*

5.2 Proof of Theorem 3.4

We rely on a Cramér-type result from general M-estimation theory.

We need to show the convergence in P_0 probability of several random sequences uniformly in $\theta \in \Theta$. In this respect it is the most convenient to rewrite (6) as

$$Q_n^1(\theta) = h_n \sum_{i=1}^{n} A_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\} + \sum_{i=1}^{n} A_{i-1}(\theta)\zeta'_{ni}, \tag{16}$$

$$\begin{aligned} Q_n^2(\theta) &= h_n \sum_{i=1}^{n} B_{i-1}(\theta)\{b_{i-1}(\beta_0)^2 - b_{i-1}(\beta)^2\}] + \sum_{i=1}^{n} B_{i-1}(\theta)\zeta''_{ni} \\ &\quad + h_n \sum_{i=1}^{n} B_{i-1}(\theta) 2\zeta'_{ni}\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\} + h_n^2 \sum_{i=1}^{n} B_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\}^2. \end{aligned} \tag{17}$$

Under the assumptions we can differentiate these quantities to obtain

$$\partial_\alpha Q_n^1(\theta) = \sum_i \partial_\alpha A_{i-1}(\theta)\zeta'_{ni} + h_n \sum_i [\partial_\alpha A_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\} - A_{i-1}(\theta)\partial_\alpha a_{i-1}(\alpha)], \tag{18}$$

$$\partial_\beta Q_n^1(\theta) = \sum_i \partial_\beta A_{i-1}(\theta)\zeta'_{ni} + h_n \sum_i \partial_\beta A_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\}, \tag{19}$$

$$\begin{aligned} \partial_\alpha Q_n^2(\theta) &= \sum_i \partial_\alpha B_{i-1}(\theta)\zeta''_{ni} + h_n \sum_i \partial_\alpha B_{i-1}(\theta)\{b_{i-1}(\beta_0)^2 - b_{i-1}(\beta)^2\} \\ &\quad + 2h_n \sum_i \zeta'_{ni}[\partial_\alpha B_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\} - B_{i-1}(\theta)\partial_\alpha a_{i-1}(\alpha)] \\ &\quad + h_n^2 \sum_i [\partial_\alpha B_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\}^2 - 2B_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\}\partial_\alpha a_{i-1}(\alpha)], \end{aligned} \tag{20}$$

$$\begin{aligned} \partial_\beta Q_n^2(\theta) &= \sum_i \partial_\beta B_{i-1}(\theta)\zeta''_{ni} + h_n \sum_i \partial_\beta B_{i-1}(\theta)\{b_{i-1}(\beta_0)^2 - b_{i-1}(\beta)^2\} \\ &\quad + 2h_n \sum_i [\partial_\beta B_{i-1}(\theta)\zeta'_{ni}\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\} - B_{i-1}(\theta)b_{i-1}(\beta)\partial_\beta b_{i-1}(\beta)] \\ &\quad + h_n^2 \sum_i \partial_\beta B_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\}^2. \end{aligned} \tag{21}$$

(Note that we presuppose that $\partial_\beta \alpha = \partial_\alpha \beta = 0$.)

Let us prove that there exist $\mathbb{R}^p$-valued measurable functions $\hat{\theta}_n = \hat{\theta}_n(X_{t_i} : i \le n)$ such that $P_0[Q_n(\hat{\theta}_n) = 0] \to 1$, and that any such sequence fulfils $\hat{\theta}_n \to^p \theta_0$. Put

$$Q(\theta) = \begin{pmatrix} Q^1(\theta) \\ Q^2(\theta) \end{pmatrix} := \begin{pmatrix} \pi_0(A(\cdot,\theta)\{a(\cdot,\alpha_0) - a(\cdot,\alpha)\}) \\ \pi_0(B(\cdot,\theta)\{b^2(\cdot,\beta_0) - b^2(\cdot,\beta)\}) \end{pmatrix}. \tag{22}$$

Under Assumption 3.3, $Q(\theta) = 0$ iff $\theta = \theta_0$, therefore,

$$\forall \epsilon > 0 \qquad \inf_{\theta \in \Theta : |\theta - \theta_0| \ge \epsilon} |Q(\theta)| > 0. \tag{23}$$

Moreover, Q is of class $C^1(\Theta)$ with the partial derivatives given by

$$\partial_\alpha Q^1(\theta) = \pi_0(\partial_\alpha A(\cdot,\theta)\{a(\cdot,\alpha_0) - a(\cdot,\alpha)\}) - \pi_0(A(\cdot,\theta)\partial_\alpha a(\cdot,\alpha)), \tag{24}$$
$$\partial_\beta Q^1(\theta) = \pi_0(\partial_\beta A(\cdot,\theta)\{a(\cdot,\alpha_0) - a(\cdot,\alpha)\}), \tag{25}$$
$$\partial_\alpha Q^2(\theta) = \pi_0(\partial_\alpha B(\cdot,\theta)\{b^2(\cdot,\beta_0) - b^2(\cdot,\beta)\}), \tag{26}$$
$$\partial_\beta Q^2(\theta) = \pi_0(\partial_\beta B(\cdot,\theta)\{b^2(\cdot,\beta_0) - b^2(\cdot,\beta)\}) - \pi(B(\cdot,\theta)\partial_\beta b^2(\cdot,\beta)). \tag{27}$$

So $\partial_\theta Q(\theta_0)$ is diagonal, and Assumption 3.3 implies that

$$\partial_\theta Q(\theta_0) = \mathrm{diag}[-\pi_0(A\partial_\alpha^\top a), -\pi_0(B\partial_\beta^\top b^2)] \text{ is nonsingular.} \tag{28}$$

In view of the general M-estimation theory combined with (23) and (28), it remains to prove (see, e.g., Yoshida [27]):

$$\sup_{\theta\in\Theta} \left| \frac{1}{T_n}\partial_\theta^l Q_n(\theta) - \partial_\theta^l Q(\theta) \right| \to^p 0 \text{ for } l \in \{0,1\}; \tag{29}$$
$$\frac{1}{T_n} Q_n(\theta_0) \to^p 0. \tag{30}$$

For clarity, from now on we focus on the proof under Assumption 3.2.1 with $\nu(\mathbb{R}) > 0$. We mention the remaining cases at the end of the proof.

The following facts are well-known and repeatedly used in the sequel without mentioning:

- Assumption 2.2 yields that $n^{-1}\sum_{i=1}^n F(X_{t_{i-1}}) \to^p \pi_0(F)$ for any smooth bounded function F with bounded derivatives (this is true as $E_0[|T_n^{-1}\int_0^{T_n} F(X_s)ds - n^{-1}\sum_{i=1}^n F(X_{t_{i-1}})|] \lesssim \sqrt{h_n}$);
- for vector-valued triangular arrays $(\chi_{ni})_{i\le n}$ with each χ_{ni} being $\mathcal{F}_{t_i}$-measurable and a random vector χ, the convergence $|\sum_{i=1}^n(\chi_{ni} - E_0[\chi_{ni}])| \to^p 0$ is implied by $\sum_{i=1}^n E_0^{i-1}[|\chi_{ni}|^2] \to^p 0$;
- if $E[|Z_t|^q] < \infty$ for some integer $q > (p \vee 2)$, $T > 1$, and c is a predictable process, then the Burkholder-Davis-Gundy inequality gives $E_0[|\int_0^T c_{s-}dZ_s|^q] \lesssim T^{q/2-1}\int_0^T E_0[|c_s|^q]ds$.

Prior to the proofs of (29) and (30), we recall two more or less well known facts.

Lemma 5.5. *Let $\{U_n(h) : h \in H\}_{n\in\mathbb{N}}$ be random fields of class C^1 defined on a bounded convex domain $H \subset \mathbb{R}^p$. Suppose the following conditions.*

1. *$U_n(h) \to^p 0$ for each $h \in H$.*
2. *Either one of the following holds true:*
   (a) $\sup_n E[\sup_h |\partial_h U_n(h)|] < \infty$;
   (b) $\sup_{n,h} E[|U_n(h)|^q] < \infty$ and $\sup_n E[\sup_h |\partial_h U_n(h)|^q] < \infty$ for a constant $q > p$ (so, $q > 1$).

Then $\sup_h |U_n(h)| \to^p 0$.

It is easy to prove that $U_n(\cdot)$ is tight with respect to the uniform metric on H under the conditions 1 and 2(a). See, e.g., Kunita [9, Section 1.4] for details concerning the tightness under the conditions 1 and 2(b).

Lemma 5.6. *Let $\{(U_{ni})_{i=1}^n\}_{n\in\mathbb{N}}$ be arrays of $\mathbb{R}^r$-valued random variables in $L^2(P)$, each U_{ni} being $\mathcal{F}_{t_i}$-measurable. Suppose the following conditions:*

1. $\sum_{i=1}^n E_0^{i-1}[U_{ni}] \to^p 0$ *and* $\sum_{i=1}^n |E_0^{i-1}[U_{ni}]|^2 \to^p 0$;
2. $\sum_{i=1}^n E_0^{i-1}[U_{ni}^{\otimes 2}] \to^p V$ *for some constant* $V \in \mathbb{R}^r \otimes \mathbb{R}^r$;
3. $\sum_{i=1}^n E_0^{i-1}[|U_{ni}|^{2+\epsilon}] \to^p 0$ *for some constant* $\epsilon > 0$.

Then $\sum_{i=1}^n U_{ni} \to^d \mathcal{N}_r(0, V)$.

See Dvoretzky [1] for details of the proof of Lemma 5.6.

Proof of (29). This follows by using Lemmas 5.1, 5.3, 5.5, and 5.6, together with the expressions (16) to (21), (22), and (24) to (27). To save space, we only prove $\sup_{\theta\in\Theta} |T_n^{-1} Q_n^1(\theta) - Q^1(\theta)| \to^p 0$, and omit the others.
We have

$$\begin{aligned}\sup_{\theta\in\Theta} |T_n^{-1} Q_n^1(\theta) - Q^1(\theta)| &\le \sup_{\theta\in\Theta} \left|\frac{1}{n}\sum_{i=1}^n A_{i-1}(\theta)\{a_{i-1}(\alpha_0) - a_{i-1}(\alpha)\} - Q^1(\theta)\right| \\ &\quad + \sup_{\theta\in\Theta}\left|\frac{1}{T_n}\sum_{i=1}^n A_{i-1}(\theta)\zeta'_{ni}\right| \\ &=: \sup_{\theta\in\Theta}|W_n^1(\theta)| + \sup_{\theta\in\Theta}|W_n^2(\theta)|.\end{aligned}$$

The θ-pointwise convergences to 0 of W_n^1 and W_n^2 follow from $E_0^{i-1}[\zeta'_{ni}] = h_n^2 R_1(X_{t_{i-1}}, \theta)$ and $E_0^{i-1}[\zeta'^2_{ni}] = h_n c_{i-1}(\beta) + h_n^2 R_1(X_{t_{i-1}}, \theta)$ (cf. Lemma 5.3), and the uniformity of the convergence of W_n^1 are implied by using Lemma 5.5 under the condition 2(a) therein. As for the uniformity of the convergence of W_n^2, we are going to apply Lemma 5.5 under the condition 2(b) therein. Fix an integer $q > (p \vee 4)$. Noting that

$$\begin{aligned}|W_n^2(\theta)|^q &\lesssim \frac{1}{T_n}\sum_{i=1}^n |A_{i-1}(\theta)|^q \int_{t_{i-1}}^{t_i} |a(X_s,\alpha_0) - a_{i-1}(\alpha_0)|^q ds \\ &\quad + \left|\frac{1}{T_n}\int_0^{T_n}\left(\sum_{i=1}^n \mathbf{1}_{(t_{i-1},t_i]}(s) A_{i-1}(\theta) b(X_s,\beta_0)\right) dZ_s\right|^q,\end{aligned}$$

we deduce, using Lemma 5.1 in part,

$$\begin{aligned}\sup_{\theta,n} E_0[|W_n^2(\theta)|^q] &\lesssim \sup_{\theta,n}\left\{\frac{1}{T_n}\sum_{i=1}^n\int_{t_{i-1}}^{t_i} E_0[E_0^{i-1}[|X_s - X_{t_{i-1}}|^q]] ds \right. \\ &\quad \left. + \frac{1}{T_n^{q/2+1}}\int_0^{T_n} E_0\left[\left(\sum_{i=1}^n \mathbf{1}_{(t_{i-1},t_i]}(s)|A_{i-1}(\theta)b(X_s,\beta_0)|\right)^q\right] ds\right\} \\ &\lesssim \sup_{\theta,n}\left(h_n + \frac{1}{T_n^{q/2+1}}\sum_{i=1}^n\int_{t_{i-1}}^{t_i} E_0[R_1(X_s,\theta)] ds\right) \lesssim 1,\end{aligned}$$

hence the first condition in 2(b) of Lemma 5.5. The second condition in 2(b) of Lemma 5.5 can be verified in a similar manner. Hence we obtain the uniformity of the convergence of W_n^2.

Proof of (30). This is automatic according to the form of $Q(\theta)$ and (29) for $l = 0$.

We now turn to the proof of the asymptotic normality; we may focus on the event $\{\omega \in \Omega : Q_n(\hat{\theta}_n) = 0\}$, whose P_0 probability tends to 1. By means of the first-order Taylor expansion $T_n^{-1}\partial_\theta Q_n(\tilde{\theta}_n)\sqrt{T_n}(\hat{\theta}_n - \theta_0) =$

$-T_n^{-1/2}Q_n(\theta_0)$, the convergence (29) for $l = 1$, and Slutsky's lemma, it suffices to prove the central limit theorem

$$\frac{1}{\sqrt{T_n}}Q_n(\theta_0) =: \sum_{i=1}^n \begin{pmatrix}\xi_{ni}^1\\ \xi_{ni}^2\end{pmatrix} \to^d \mathcal{N}_p\left(0, \begin{pmatrix} \pi_0(A^{\otimes 2}b^2) & \nu_3\pi_0(AB^\top b^3)\\ \text{Sym.} & \nu_4\pi_0(B^{\otimes 2}b^4)\end{pmatrix}\right), \tag{31}$$

where $\xi_{ni}^1 \in \mathbb{R}^{p_\alpha}$ and $\xi_{ni}^2 \in \mathbb{R}^{p_\beta}$. From Lemma 5.3 we easily get $E_0^{i-1}[\xi_{ni}^j] = h_n^{3/2}R_1(X_{t_{i-1}},\theta)$ and $E_0^{i-1}[|\xi_{ni}^j|^q] = n^{-1}T_n^{1-q/2}R_1(X_{t_{i-1}},\theta)$ for $j \in \{1,2\}$ and $q > 2$, so that, in view of Lemma 5.6 it remains to prove the convergence of the predictable quadratic covariation matrix. As a matter of fact, Lemma 5.3 leads to the desired convergence:

$$\begin{aligned}
\sum_{i=1}^n E_0^{i-1}[\xi_{ni}^{1\,\otimes 2}] &= \frac{1}{n}\sum_{i=1}^n A_{i-1}^{\otimes 2}(\theta_0)\frac{1}{h_n}E_0^{i-1}[\zeta_{ni}'^{\,2}] + O_p(h_n)\\
&= \pi_0(A^{\otimes 2}b^2) + o_p(1),\\
\sum_{i=1}^n E_0^{i-1}[\xi_{ni}^1\xi_{ni}^{2\top}] &= \frac{1}{n}\sum_{i=1}^n A_{i-1}(\theta_0)B_{i-1}^\top(\theta_0)\frac{1}{h_n}E_0^{i-1}[\zeta_{ni}'^{\,3}] + O_p(h_n^2)\\
&= \nu_3\pi_0(AB^\top b^3) + o_p(1),\\
\sum_{i=1}^n E_0^{i-1}[\xi_{ni}^{2\,\otimes 2}] &= \frac{1}{n}\sum_{i=1}^n B_{i-1}^{\otimes 2}(\theta_0)\frac{1}{h_n}E_0^{i-1}[\zeta_{ni}'^{\,4}] + O_p(h_n)\\
&= \nu_4\pi_0(B^{\otimes 2}b^4) + o_p(1),
\end{aligned}$$

leading to (31). Thus we have proved Theorem 3.4 under Assumption 3.2.1 with $\nu(\mathbb{R}) \in (0,\infty]$.

Concerning the consistent estimators of V_0' and V_0'', the case of (7) is obvious from Assumption 2.2 and Lemma 5.5. Turning to the case (8), we only mention $\nu_4\hat{V}_{22,n}$. Put $\nu_4\hat{V}_{22,n} = \sum_{i=1}^n \eta_{ni}$. By means of Lemma 5.3 we readily get $\sum_{i=1}^n E_0^{i-1}[\eta_{ni}] = \nu_4 V_{22} + o_p(1)$. Moreover, since $\nu_8 < \infty$ under Assumption 2.3, $\sum_{i=1}^n E_0^{i-1}[\eta_{ni}^{\otimes 2}] = T_n^{-1}\{\nu_8 n^{-1}\sum_{i=1}^n [(B^{\otimes 2})^{\otimes 2}b^8](X_{t_{i-1}},\theta_0) + o_p(1)\} = o_p(1)$ as was to be shown.

Finally, let us briefly mention the remaining cases in proving Theorem 3.4. The proof for cases where $\nu(\mathbb{R}) > 0$ under Assumption 3.2.2 can be achieved in the same way as above except that we should replace all the appearing R_1 by R_2. As for the diffusion cases under Assumption 3.2, the only difference is, of course, that we need to use Lemma 5.2 instead of Lemma 5.3, together with the different norming of Q_n. We may omit the details, for it is well known in the literature under Assumption 3.2.2 (cf. Kessler [5] and Sørensen [19]), and the proof under Assumption 3.2.1 can be achieved all without distinction by making use of Lemma 5.1 with g bounded.

5.3 Proof of Theorem 4.1

Put $\epsilon_{ni} = \epsilon_{ni}(\theta_0)$, and define

$$H_n^{(k)} := \frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^k,\quad \hat{H}_n^{(k)} := \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_{ni}^k,\quad \hat{M}_n^{(k,j)} := \frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{k-j}(\hat{\epsilon}_{ni}-\epsilon_{ni})^j.$$

It is convenient to expand $\hat{\Psi}_n^{(k)}$ and $\hat{H}_n^{(k)}$ as follows: for $k \in \mathbb{N}$,

$$\hat{\Psi}_n^{(k)} = \hat{H}_n^{(k)} - k\hat{H}_n^{(1)}\hat{H}_n^{(k-1)} + \sum_{j=2}^k \binom{k}{j}(-\hat{H}_n^{(1)})^j\hat{H}_n^{(k-j)}, \tag{32}$$

$$\hat{H}_n^{(k)} = H_n^{(k)} + k\hat{M}_n^{(k,1)} + \sum_{j=2}^k \binom{k}{j}\hat{M}_n^{(k,j)}. \tag{33}$$

Once again, we will prove the claims only under Assumption 2.1.1, as the case of Assumption 2.1.2 can be treated in the same way under the condition $\sup_{t\in\mathbb{R}_+} E_0[|X_t|^q] < \infty$ for every $q > 0$.

5.3.1 Under the null hypothesis

First we prove $\mathcal{T}_n \to^d \chi^2(2)$ under $\mathcal{H}_0$. We write $\bar{\alpha}_n = \sqrt{T_n}(\hat{\alpha}_n - \alpha_0)$ and $\bar{\beta}'_n = \sqrt{n}(\hat{\beta}_n - \beta_0)$, both being $O_p(1)$. In what follows we abbreviate $b_{i-1}(\hat{\beta}_n)$ to $\hat{b}_{i-1}$, and also, $R_1(X_{t_{i-1}}, \theta)$ to R_{i-1}.

First we note the following direct consequences of Lemma 5.2:

$$H_n^{(k)} \to^p \rho_k \quad \text{for each } k \in \mathbb{N}, \text{and especially} \quad H_n^{(1)} = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{34}$$

The next lemma shows that, thanks to the centering factor $\bar{\hat{\epsilon}}_n$ in the summands of $\hat{\Psi}_n^{(k)}$, the effect of "plugging-in $\hat{\alpha}_n$" disappears from the expressions up to order $o_p(1/\sqrt{n})$.

Lemma 5.7. *For each $k \in \mathbb{N}$ we have*

$$\hat{\Psi}_n^{(k)} = (H_n^{(k)} - kH_n^{(1)}H_n^{(k-1)}) - \frac{k\rho_k}{\sqrt{n}}\gamma_n^{(b)}\bar{\beta}'_n + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{35}$$

where $\gamma_n^{(b)} := n^{-1}\sum_{i=1}^n (\partial_\beta^\top b_{i-1}/b_{i-1})(\beta_0) = O_p(1)$.

Proof. Applying the Taylor expansion under Assumption 3.2, we obtain

$$\begin{aligned}
&\sqrt{n}(\hat{\epsilon}_{ni} - \epsilon_{ni}) \\
&= \hat{b}_{i-1}^{-1}\Bigg[- \epsilon_{ni}\left\{\partial_\beta b_{i-1}(\beta_0)\bar{\beta}'_n + \frac{1}{\sqrt{n}}\left(\int_0^1 u \int_0^1 \partial_\beta^2 b_{i-1}(\beta_0 + uv(\hat{\beta}_n - \beta_0))dvdu\right)[\bar{\beta}_n'^{\otimes 2}]\right\} \\
&\quad - \left\{\partial_\alpha a_{i-1}(\alpha_0)\bar{\alpha}_n + \frac{1}{\sqrt{T_n}}\left(\int_0^1 u \int_0^1 \partial_\alpha^2 a_{i-1}(\alpha_0 + uv(\hat{\alpha}_n - \alpha_0))dvdu\right)[\bar{\alpha}_n^{\otimes 2}]\right\}\Bigg] \\
&= \hat{b}_{i-1}^{-1}\left\{ - \epsilon_{ni}\partial_\beta^\top b_{i-1}(\beta_0)\bar{\beta}'_n - \partial_\alpha a_{i-1}(\alpha_0)\bar{\alpha}_n + \epsilon_{ni}\frac{1}{\sqrt{n}}R_{i-1}[\bar{\beta}_n'^{\otimes 2}] + \frac{1}{\sqrt{T_n}}R_{i-1}[\bar{\alpha}_n^{\otimes 2}]\right\},
\end{aligned} \tag{36}$$

where $G[\bar{\alpha}_n^{\otimes 2}] := \bar{\alpha}_n^\top G\bar{\alpha}_n$ for $G \in \mathbb{R}^{p_\alpha} \otimes \mathbb{R}^{p_\alpha}$, with a similar notation for $G[\bar{\beta}_n'^{\otimes 2}]$. In the same way,

$$\hat{b}_{i-1}^{-1} = b_{i-1}^{-1} + \frac{1}{\sqrt{n}}R_{i-1}\bar{\beta}'_n. \tag{37}$$

It follows from (34), (36), and (37) that for $k \in \mathbb{N}$

$$\begin{aligned}
\sqrt{n}\hat{M}_n^{(k,1)} &= -\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^k\left(\frac{\partial_\beta^\top b_{i-1}}{b_{i-1}}\right)(\beta_0)\bar{\beta}'_n - \frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{k-1}\left(\frac{\partial_\alpha^\top a_{i-1}}{b_{i-1}}\right)(\theta_0)\bar{\alpha}_n + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= -\rho_k\frac{1}{n}\sum_{i=1}^n\left(\frac{\partial_\beta^\top b_{i-1}}{b_{i-1}}\right)(\beta_0)\bar{\beta}'_n - \rho_{k-1}\frac{1}{n}\sum_{i=1}^n\left(\frac{\partial_\alpha^\top a_{i-1}}{b_{i-1}}\right)(\theta_0)\bar{\alpha}_n + O_p\left(\frac{1}{\sqrt{n}}\right) \\
&= -\rho_k\gamma_n^{(b)}\bar{\beta}'_n - \rho_{k-1}\gamma_n^{(a)}\bar{\alpha}_n + O_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{say.}
\end{aligned} \tag{38}$$

In particular, (38) implies that $\hat{M}_n^{(k,1)} = O_p(1/\sqrt{n})$. In view of (36), we see that

$$\frac{1}{n}\sum_{i=1}^n |\sqrt{n}(\hat{\epsilon}_{ni} - \epsilon_{ni})|^l = O_p(1)$$

for any $l \in \mathbb{N}$ under Assumption 3.2, so that for $j \geq 2$

$$|\sqrt{n}\hat{M}_n^{(k,j)}| = n^{(1-j)/2}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{k-j}\{\sqrt{n}(\hat{\epsilon}_{ni} - \epsilon_{ni})\}^j\right| = O_p(n^{(1-j)/2}) = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Hence we have obtained that for $k \in \mathbb{N}$:

$$M_n^{(k,j)} = \begin{cases} O_p(1/\sqrt{n}), & j = 1; \\ O_p(1/n), & j \geq 2. \end{cases} \tag{39}$$

Combining (33), (34), and (39) now gives $\hat{H}_n^{(k)} = O_p(1)$, and especially $\hat{H}_n^{(1)} = O_p(1/\sqrt{n})$. Therefore (32) and (33) actually become $\hat{\Psi}_n^{(k)} = \hat{H}_n^{(k)} - k\hat{H}_n^{(1)}\hat{H}_n^{(k-1)} + O_p(1/n)$ and $\hat{H}_n^{(k)} = H_n^{(k)} + k\hat{M}_n^{(k,1)} + O_p(1/n)$, respectively, from which we get

$$\hat{\Psi}_n^{(k)} = (H_n^{(k)} - kH_n^{(1)}H_n^{(k-1)}) + k\hat{F}_n^{(k)} + O_p\left(\frac{1}{\sqrt{n}}\right) \tag{40}$$

for each $k \in \mathbb{N}$, where $\hat{F}_n^{(k)} := \hat{M}_n^{(k,1)} - H_n^{(k-1)}\hat{M}_n^{(1,1)}$. Now substituting (38) in (40) and then rearranging the resulting terms, we arrive at (35). □

Put $C_n^{(k)} = n^{-1}\sum_{i=1}^n E_0^{i-1}[\epsilon_{ni}^k]$ and $\tilde{H}_n^{(k)} = H_n^{(k)} - C_n^{(k)}$; note that $C_n^{(k)} \to^p \rho_k$. As seen in the next lemma, the self-normalizing factors $(\hat{\Psi}_n^{(2)})^{k/2}$ in the definitions of $\hat{\Phi}_n^{(k)}$ disable the effect of "plugging-in $\hat{\beta}_n$" up to order $o_p(1/\sqrt{n})$.

Lemma 5.8. *For each $k \in \mathbb{N}$ we have*

$$\hat{\Phi}_n^{(k)} - C_n^{(k)} = \left(\tilde{H}_n^{(k)} - k\rho_{k-1}\tilde{H}_n^{(1)} - \frac{k}{2}\rho_k\tilde{H}_n^{(2)}\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{41}$$

Proof. In view of Lemma 5.7, we see that

$$\hat{\Psi}_n^{(k)} = C_n^{(k)} + \frac{1}{\sqrt{n}}\{\sqrt{n}\tilde{H}_n^{(k)} - k\rho_{k-1}\sqrt{n}\tilde{H}_n^{(1)} - k\rho_k\gamma_n^{(b)}\bar{\beta}_n'\} + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{42}$$

This in particular yields that for $k \geq 2$

$$\begin{aligned}(\hat{\Psi}_n^{(2)})^{k/2} &= \left\{1 + \frac{1}{\sqrt{n}}(\sqrt{n}\tilde{H}_n^{(2)} - 2\gamma_n^{(b)}\bar{\beta}_n') + o_p\left(\frac{1}{\sqrt{n}}\right)\right\}^{k/2} \\ &= 1 + \frac{k}{2\sqrt{n}}(\sqrt{n}\tilde{H}_n^{(2)} - 2\gamma_n^{(b)}\bar{\beta}_n') + o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{43}$$

since $C_n^{(2)} = 1 + O_p(h_n) = 1 + o_p(1/\sqrt{n})$ under $nh_n^2 \to 0$. Thus expanding the fraction $\hat{\Psi}_n^{(k)}/(\hat{\Psi}_n^{(2)})^{k/2}$ by using (42) and (43), we get

$$\begin{aligned}\hat{\Phi}_n^{(k)} - C_n^{(k)} = \frac{1}{\sqrt{n}}\Big\{&\sqrt{n}\tilde{H}_n^{(k)} - k\rho_{k-1}\sqrt{n}\tilde{H}_n^{(1)} - k\rho_k\gamma_n^{(b)}\bar{\beta}_n' \\ &- \frac{k}{2}C_n^{(k)}(\sqrt{n}\tilde{H}_n^{(2)} - 2\gamma_n^{(b)}\bar{\beta}_n')\Big\} + o_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Using $\tilde{H}_n^{(2)} = O_p(1/\sqrt{n})$ (apply Lemma 5.6) and $C_n^{(k)} = \rho_n + o_p(1)$, it is obvious that the right-hand side being identical to that of (41). □

Finally, we apply Lemma 5.6 to the expressions (41) for $k = 3$ and 4.

Lemma 5.9. *We have*

$$\sqrt{n}\begin{pmatrix}\hat{\Phi}_n^{(3)} - \dfrac{3\sqrt{h_n}}{n}\displaystyle\sum_{i=1}^n \partial_x b(X_{t_{i-1}}, \hat{\beta}_n) \\ \hat{\Phi}_n^{(4)} - 3\end{pmatrix} \to^d \mathcal{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}6 & 0\\ 0 & 24\end{pmatrix}\right)$$

Proof. By Lemma 5.2 and Taylor's expansion of $\beta_0 \mapsto \partial_x b(X_{t_{i-1}}, \beta_0)$ around $\hat{\beta}_n$, we have

$$\begin{aligned}
C_n^{(3)} &= \frac{1}{n}\sum_{i=1}^{n}\{3\sqrt{h_n}\partial_x b(X_{t_{i-1}}, \beta_0) + h_n^{3/2}R_1(X_{t_{i-1}}, \theta)\} \\
&= \frac{3\sqrt{h_n}}{n}\sum_{i=1}^{n}\partial_x b(X_{t_{i-1}}, \hat{\beta}_n) + O_p\left(\left(|\bar{\beta}_n'|\sqrt{\frac{h_n}{n}}\right) \vee h_n^{3/2}\right) \\
&= \frac{3\sqrt{h_n}}{n}\sum_{i=1}^{n}\partial_x b(X_{t_{i-1}}, \hat{\beta}_n) + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Hence it follows from (41) that

$$\begin{aligned}
&\sqrt{n}\left\{\hat{\Phi}_n^{(3)} - \frac{3\sqrt{h_n}}{n}\sum_{i=1}^{n}\partial_x b(X_{t_{i-1}}, \hat{\beta}_n)\right\} \\
&= \sum_{i=1}^{n}\frac{1}{\sqrt{n}}\left\{(\epsilon_{ni}^3 - E_0^{i-1}[\epsilon_{ni}^3]) - 3(\epsilon_{ni} - E_0^{i-1}[\epsilon_{ni}])\right\} + o_p(1).
\end{aligned} \tag{44}$$

We similarly get $C_n^{(4)} = 3 + O_p(h_n) = 3 + o_p(1/\sqrt{n})$, so that

$$\sqrt{n}(\hat{\Phi}_n^{(4)} - 3) = \sum_{i=1}^{n}\frac{1}{\sqrt{n}}\left\{(\epsilon_{ni}^4 - E_0^{i-1}[\epsilon_{ni}^4]) - 6(\epsilon_{ni}^2 - E_0^{i-1}[\epsilon_{ni}^2])\right\} + o_p(1). \tag{45}$$

We deduce the claim by applying Lemma 5.6 together with Lemma 5.2 to the first terms of the right-hand sides of (44) and (45). □

Lemma 5.9 combined with the continuous mapping theorem completes the proof of Theorem 4.1 under $\mathcal{H}_0$.

5.3.2 Under the alternative hypothesis

Next we prove that $P_0[\mathcal{T}_n > K] \to 1$ for every $K > 0$ under $\mathcal{H}_1$. Write $\bar{\beta}_n'' = \sqrt{T_n}(\hat{\beta}_n - \beta_0)$, which has the Gaussian weak limit as specified in Theorem 3.4. Fix any $K > 0$ in the sequel. According to the definition of $\mathcal{T}_n$, it suffices to show that

$$P_0[|\Lambda_n^{(3)}| > K] \to 1, \tag{46}$$

where $\Lambda_n^{(3)} = \sqrt{n}\{\hat{\Phi}_n^{(3)} - 3n^{-1}\sqrt{h_n}\sum_{i=1}^{n}\partial_x b(X_{t_{i-1}}, \hat{\beta}_n)\}$.

Using (32), we rewrite $\Lambda_n^{(3)}$ as

$$\begin{aligned}
\Lambda_n^{(3)} = (\hat{\Psi}_n^{(2)})^{-3/2}\frac{1}{h_n}\bigg\{&\sqrt{nh_n^2}\hat{H}_n^{(3)} - 3\sqrt{nh_n^2}\hat{H}_n^{(1)}\hat{H}_n^{(2)} + 2\sqrt{nh_n^2}\hat{H}_n^{(2)} \\
&- (\hat{\Psi}_n^{(2)})^{3/2}\sqrt{nh_n^3}\frac{1}{n}\sum_{i=1}^{n}\partial_x b(X_{t_{i-1}}, \hat{\beta}_n)\bigg\}.
\end{aligned} \tag{47}$$

It is convenient to note $\Delta_i X - h_n a_{i-1}(\hat{\alpha}_n) = \zeta_{ni}' + \delta_{ni}\sqrt{h_n/n}$, where $\delta_{ni} = R_{i-1}\bar{\alpha}_n$. Then

$$\hat{H}_n^{(k)} = h_n^{-k/2}\left\{\sum_{j=0}^{k-1}\binom{k}{j}n^{-j/2}h_n^{1+j/2}\frac{1}{T_n}\sum_{i=1}^{n}\hat{b}_{i-1}^{-k}\zeta_{ni}'^{\,k-j}\delta_{ni}^j + n^{-k/2}\frac{1}{n}\sum_{i=1}^{n}\hat{b}_{i-1}^{-k}\delta_{ni}^k\right\}, \tag{48}$$

from which combined with Lemma 5.3 we deduce

$$\begin{aligned}
\hat{H}_n^{(1)} &= \sqrt{h_n}\frac{1}{T_n}\sum_{i=1}^{n}\hat{b}_{i-1}^{-1}\zeta_{ni}' + O_p\left(\frac{1}{\sqrt{n}}\right) = o_p(1), \\
\hat{H}_n^{(2)} &= \frac{1}{T_n}\sum_{i=1}^{n}\hat{b}_{i-1}^{-2}\zeta_{ni}'^{\,2} + 2\sqrt{\frac{h_n}{n}}\frac{1}{T_n}\sum_{i=1}^{n}\hat{b}_{i-1}^{-2}\zeta_{ni}'\delta_{ni} + O_p\left(\frac{1}{n}\right) = 1 + o_p(1).
\end{aligned}$$

(We used $\hat{b}_{i-1}^{-2} = b_{i-1}(\beta_0)^{-2} + T_n^{-1/2} R_{i-1} \bar{\beta}_n''$ for the latter.) Hence we also have $\hat{\Psi}_n^{(2)} = \hat{H}_n^{(2)} - (\hat{H}_n^{(1)})^2 = 1 + o_p(1)$, so that (47) becomes

$$\Lambda_n^{(3)} = \frac{1}{h_n}(1 + o_p(1)) \left\{ \sqrt{n h_n^2} \hat{H}_n^{(3)} + o_p(1) \right\}.$$

We are going to prove that $\sqrt{n h_n^2} \hat{H}_n^{(3)}$ does not tend in P_0 probability to 0, from which (46) follows.

Using (48) with Lemma 5.3 as before, it is easy to see that

$$\begin{aligned}
\sqrt{n h_n^2} \hat{H}_n^{(3)} &:= \sum_{i=1}^n \frac{1}{\sqrt{T_n}} \hat{b}_{i-1}^{-3} \{ \zeta_{ni}'^{\,3} - h_n \nu_3 b_{i-1}^3(\beta_0) \} + \nu_3 \sqrt{T_n} \frac{1}{n} \sum_{i=1}^n \left(\frac{b_{i-1}(\beta_0)}{\hat{b}_{i-1}} \right)^3 + o_p(1) \\
&= \sum_{i=1}^n \frac{1}{\sqrt{T_n}} b_{i-1}^{-3}(\beta_0) \{ \zeta_{ni}^3 - h_n \nu_3 b_{i-1}^3(\beta_0) \} + \nu_3 \sqrt{T_n} \frac{1}{n} \sum_{i=1}^n \left(\frac{b_{i-1}(\beta_0)}{\hat{b}_{i-1}} \right)^3 + o_p(1) \\
&=: I_{1n} + I_{2n} + o_p(1).
\end{aligned}$$

Write $I_{1n} = \sum_{i=1}^n \chi_{ni}$, then Lemma 5.3 gives that $E_0^{i-1}[\chi_{ni}] = T_n^{-1/2} h_n^2 R_{i-1}$. Also, for each $\delta \in (0, q-2)$,

$$\begin{aligned}
\sum_{i=1}^n E_0^{i-1}[|\chi_{ni}|^{2+\delta}] &\lesssim T_n^{-\delta/2} \frac{1}{T_n} \sum_{i=1}^n E_0^{i-1}[|\zeta_{ni}'^{\,3} - h_n \nu_3 b_{i-1}(\beta_0)^3|^{2+\delta}] \\
&\lesssim T_n^{-\delta/2} \frac{1}{T_n} \sum_{i=1}^n (E_0^{i-1}[|\Delta_i X|^{6+3\delta}] + h_n^{6+3\delta} + \nu_3^{2+\delta} h_n^{2+\delta}) \\
&\lesssim T_n^{-\delta/2} \frac{1}{n} \sum_{i=1}^n (1 + h_n^{5+3\delta} + \nu_3^{2+\delta} h_n^{1+\delta}) = O_p(T_n^{-\delta/2}) = o_p(1).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\sum_{i=1}^n E_0^{i-1}[|\chi_{ni}|^2] &= \frac{1}{T_n} \sum_{i=1}^n b_{i-1}(\beta_0)^{-6} \{ E_0^{i-1}[\zeta_{ni}'^{\,6}] - 2 h_n \nu_3 b_{i-1}(\beta_0)^3 E_0^{i-1}[\zeta_{ni}'^{\,3}] + h_n^2 \nu_3^2 b_{i-1}(\beta_0)^6 \} \\
&= \frac{1}{n} \sum_{i=1}^n (\nu_6 - h_n \nu_3^2) + o_p(h_n) \to^p \nu_6.
\end{aligned}$$

Thus Lemma 5.6 yields that $I_{1n} \to^d \mathcal{N}_1(0, \nu_6)$. On the other hand, we have $I_{2n} \equiv 0$ if $\nu_3 = 0$, but otherwise the sequence (I_{2n}) is not stochastically bounded since $I_{2n} = \nu_3 \sqrt{T_n}(1 + o_p(1))$. The proof of Theorem 4.1 is thus complete.

Acknowledgement. The author sincerely extends his thanks to Professor Alexey M. Kulik, who gave him a better understanding of a result in Kulik [6]. This work was partly supported by Grant-in-Aid for Young Scientists (B) 20740061, Japan, and Cooperative Research Program of the Institute of Statistical Mathematics.

References

[1] Dvoretzky, A. (1977), Asymptotic normality of sums of dependent random vectors. *Multivariate analysis, IV* (Proc. Fourth Internat. Sympos., Dayton, Ohio, 1975), pp. 23–34. North-Holland, Amsterdam.

[2] Henze, N. (2002), Invariant tests for multivariate normality: a critical review. *Statist. Papers* **43**, 467–506.

[3] Heyde, C. C. (1997), *Quasi-Likelihood and its Application. A General Approach to Optimal Parameter Estimation.* Springer-Verlag, New York.

[4] Heyde, C. C. and Leonenko, N. N. (2005), Student processes. *Adv. in Appl. Probab.* **37**, 342–365.

[5] Kessler, M. (1997), Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211–229.

[6] Kulik, A. M., (2009), Exponential ergodicity of the solutions to SDE's with a jump noise. *Stochastic Process. Appl.* **119**, 602–632.

[7] Kulik, A. M., (2009), Personal communications.

[8] Kulperger, R. and Yu, H. (2005), High moment partial sum processes of residuals in GARCH models and their applications. *Ann. Statist.* **33**, 2395–2422.

[9] Kunita, H. (1990), *Stochastic Flows and Stochastic Differential Equations*. Cambridge University Press, Cambridge.

[10] Lee, S. and Masuda, H. (2008), Jarque-Bera normality test for the driving Lévy process of a discretely observed univariate SDE. *Stat. Inference Stoch. Process.* **13**, 147–161.

[11] Masuda, H. (2005), Simple estimators for parametric Markovian trend of ergodic processes based on sampled data. *J. Japan Statist. Soc.* **35**, 147–170.

[12] Masuda, H. (2007), Ergodicity and exponential β-mixing bound for multidimensional diffusions with jumps. *Stochastic Processes Appl.* **117**, 35–56. [Erratum to this paper: (2009) *Stochastic Processes Appl.* **119**, 676–678.]

[13] Masuda, H. (2008), On stability of diffusions with compound-Poisson jumps. *Bull. Inform. Cybernet.* **40**, 61–74.

[14] Masuda, H. (2010), Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test. MI Preprint Series, Kyushu University, 2010-21. Available at http://gcoe-mi.jp/temp/publish/bdfa1e74d21a7ee1a912e6b8877a576b.pdf

[15] Protter, P. E. (2004), *Stochastic Integration and Differential Equations. Second Edition.* Springer-Verlag, Berlin.

[16] Roberts, G. O. and Tweedie, R. L. (1996), Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**, 341–363.

[17] Sato, K. (1999), *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge.

[18] Shimizu, Y. and Yoshida, N. (2006), Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat. Inference Stoch. Process.* **9**, 227–277.

[19] Sørensen, M. (2008), Efficient estimation for ergodic diffusions sampled at high frequency. Preprint.

[20] Stramer, O. and Tweedie, R. L. (1999), Langevin-type models. I. Diffusions with given stationary distributions and their discretizations. *Methodol. Comput. Appl. Probab.* **1**, 283–306.

[21] Uchida, M. and Yoshida, N. Estimation for misspecified ergodic diffusion processes from discrete observations. To appear in *ESAIM Probab. Statist.*

[22] Veretennikov, A. Y. (1997), On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl.* **70**, 115–127.

[23] Veretennikov, A. Y. (2006), On lower bounds for mixing coefficients of Markov diffusions. *From Stochastic Calculus to Mathematical Finance*, 623–633, Springer, Berlin.

[24] Wedderburn, R. W. M. (1974), Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439–447.

[25] Yoshida, N. (1992), Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41**, 220–242.

[26] Yoshida, N. (2005), Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. To appear in *Ann. Inst. Statist. Math.*

[27] Yoshida, N. *Stochastic Analysis and Statistics*. Springer. In preparation.

[28] Yu, H. (2007), High moment partial sum processes of residuals in ARMA models and their applications. *J. Time Ser. Anal.* **28**, 72–91.